

MOCHA: A Multi-site One-shot Lossless Cox Hazard Analysis

Yuqing Lei^{1,2}, Lu Li^{2,3}, Huilin Tang^{1,2}, Yudong Wang^{1,2}, Tingyin Wang^{2,3}, Siqi Chen^{2,3}, Yue Wu^{1,2}, Bingyu Zhang^{2,3}, Jie Hu^{1,2}, Huiyuan Wang^{1,2}, Jingna Feng⁴, Zifang Kong⁵, Chengrong Wang⁶, Yujia Zhou⁷, Ying Lu⁸, Anastassia Amaro⁹, Joshua Kueiyu Lin¹⁰, Chongliang Luo¹¹, Mei Liu⁶, Yulun Liu⁵, Cui Tao⁴, Hua Xu⁷, Bradley Malin^{12*}, Jingmei Qiu^{13*}, Yiwen Lu^{2,3*}, David A. Asch¹⁴, Yong Chen^{1-3,15-17}

¹ Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

² The Center for Health AI and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, PA, USA

³ The Graduate Group in Applied Mathematics and Computational Science, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA

⁴ Department of Artificial Intelligence and Informatics (AI&I), Mayo Clinic, Rochester, MN, USA

⁵ Department of Public Health, School of Public Health, The University of Texas Southwestern Medical Center, Dallas, TX, USA

⁶ Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, USA

⁷ Department of Biomedical Informatics and Data Science, School of Medicine, Yale University, New Haven, CT, USA

⁸ Department of Biomedical Data Science, School of Medicine, Stanford University, Stanford, CA, USA

⁹ Clinical Medicine in Division of Endocrinology, Diabetes and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

¹⁰ Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

¹¹ Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA

¹² Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

¹³ Department of Mathematical Sciences, University of Delaware, Newark, DE, USA

¹⁴ Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

¹⁵ Leonard Davis Institute of Health Economics, Philadelphia, PA, USA

¹⁶ Penn Medicine Center for Evidence-based Practice (CEP), Philadelphia, PA, USA

¹⁷ Penn Institute for Biomedical Informatics (IBI), Philadelphia, PA, USA

Background

Target trial emulation has become a core framework for generating real-world evidence by aligning observational analyses with the principles of randomized trials [1,2]. However, many applications focus on average treatment effects, even though treatment benefit and harm may differ across clinically meaningful subgroups [3]. For time-to-event outcomes, estimating heterogeneous treatment effects often requires data from multiple health systems, because any single site may have limited sample size, sparse events, or restricted subgroup representation. Direct pooling of patient-level data is often infeasible because of privacy, governance, and institutional constraints. Existing distributed approaches address part of this challenge, but they often require multiple communication rounds or approximate rather than reproduce the centralized pooled Cox analysis [4-6]. We developed MOCHA, a Multi-site One-shot Lossless Cox Hazard Analysis framework, to support one-shot target trial emulation for heterogeneous treatment-effect estimation without sharing individual-level data.

Methods

MOCHA uses a harmonized study protocol across participating sites, including common eligibility criteria, covariates, subgroup definitions, outcomes, propensity score models, and Cox model specifications. Each

site estimates propensity scores to balance observed confounders [11], creates propensity score strata, and evaluates its stratified Cox risk function over a prespecified Chebyshev grid spanning treatment, modifier, and interaction parameters. The resulting evaluation tensor is compressed into a low-rank tensor-train representation to reduce communication burden [9,10]. Each site transmits this compact summary once to the coordinating center, which reconstructs the pooled risk function and estimates subgroup-specific treatment effects with controllable approximation error. To address residual systematic bias, MOCHA incorporates negative control outcomes through empirical calibration [7,8]. Figure 1 summarizes the MOCHA workflow, including protocol harmonization, local risk-function evaluation, tensor-train compression, one-shot aggregation, pooled-equivalent Cox estimation, and negative-control-outcome calibration.

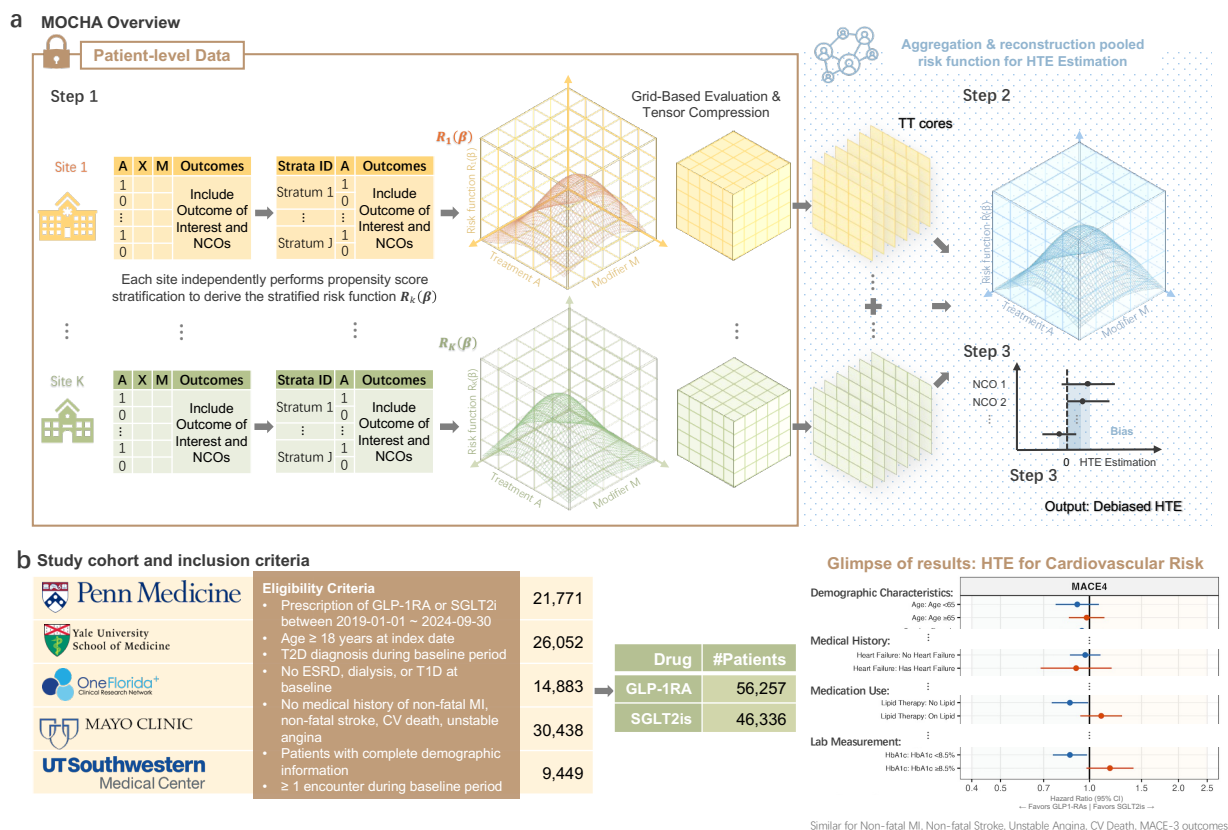


Figure 1. Overview of the MOCHA framework and decentralized real-world application.

MOCHA enables federated heterogeneous treatment-effect estimation for time-to-event outcomes through protocol harmonization, local risk-function evaluation, tensor-train compression, one-shot aggregation, pooled-equivalent Cox estimation, and negative-control-outcome calibration.

Results

In internal validation, MOCHA reproduced centralized pooled Cox estimates with high numerical fidelity. Point estimates aligned with the pooled gold standard, with a maximum absolute difference of 4×10^{-4} , and standard errors were similarly preserved, with a maximum absolute difference below 2×10^{-3} . Fixed-effect meta-analysis and ODACH showed larger discrepancies, particularly for modifier and interaction terms, where sparse site-specific events can make local estimation unstable [5,6]. In the real-world application, MOCHA was implemented across five U.S. health systems: University of Pennsylvania Health

System, Yale New Haven Health System, Mayo Clinic, OneFlorida+ Clinical Research Network, and UT Southwestern Medical Center. The analysis included 102,593 adults with type 2 diabetes initiating glucagon-like peptide-1 receptor agonists (GLP-1RAs) or sodium-glucose cotransporter 2 inhibitors (SGLT2is). We evaluated 19 prespecified potential effect modifiers across six cardiovascular outcomes, including non-fatal myocardial infarction, non-fatal stroke, unstable angina, cardiovascular death, MACE-3, and MACE-4. Consistent with prior comparative effectiveness studies of these drug classes [12,13], subgroup-specific hazard ratios were generally centered near 1 with overlapping confidence intervals, suggesting limited treatment-effect heterogeneity. For MACE-3 and MACE-4, most estimates modestly favored GLP-1RAs. Among patients with baseline HbA1c <8.5%, GLP-1RAs were associated with lower risk of MACE-3 and MACE-4 (hazard ratio 0.85, 95% confidence interval 0.74-0.97). Among patients not receiving lipid-lowering therapy at baseline, GLP-1RAs were associated with lower risk of MACE-3 (hazard ratio 0.86, 95% confidence interval 0.74-0.99), MACE-4 (hazard ratio 0.85, 95% confidence interval 0.74-0.99), and cardiovascular death (hazard ratio 0.70, 95% confidence interval 0.50-0.98). Empirical calibration using 85 negative control outcomes widened confidence intervals and attenuated one boundary association to non-significance, while preserving the remaining associations that were robust to systematic-error adjustment.

Conclusion

MOCHA enables one-shot, pooled-equivalent Cox analysis for federated target trial emulation of heterogeneous treatment effects. In validation, it reproduced centralized pooled inference more closely than meta-analysis or ODACH while avoiding patient-level data sharing. In a multi-system study of adults with type 2 diabetes, MOCHA provided calibrated subgroup-specific estimates for cardiovascular outcomes comparing GLP-1RAs with SGLT2is. These results support MOCHA as a scalable approach for multi-institutional real-world evidence generation.

References

1. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183:758-764.
2. Hernan MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. *JAMA*. 2022;328:2446-2447.
3. Angus DC, Chang CH. Heterogeneity of Treatment Effect: Estimating How the Effects of Interventions Vary Across Individuals. *JAMA*. 2021;326:2312-2313.
4. Hripcsak G, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform*. 2015;216:574-578.
5. Lu CL, et al. WebDISCO: a web service for distributed Cox model learning without patient-level data sharing. *J Am Med Inform Assoc*. 2015;22:1212-1219.
6. Luo C, et al. ODACH: a one-shot distributed algorithm for Cox model with heterogeneous multi-center data. *Sci Rep*. 2022;12:6627.
7. Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med*. 2014;33:209-218.
8. Schuemie MJ, Hripcsak G, Ryan PB, Madigan D, Suchard MA. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc Natl Acad Sci U S A*. 2018;115:2571-2577.
9. Oseledets I, Tyrtysnikov E. TT-cross approximation for multidimensional arrays. *Linear Algebra Appl*. 2010;432:70-88.
10. Oseledets IV. Tensor-Train Decomposition. *SIAM J Sci Comput*. 2011;33:2295-2317.
11. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in

observational studies. *Multivariate Behav Res.* 2011;46:399-424.

12. Patorno E, et al. Sodium-Glucose Cotransporter-2 Inhibitors Versus Glucagon-like Peptide-1 Receptor Agonists and the Risk for Cardiovascular Outcomes in Routine Care Patients With Diabetes Across Categories of Cardiovascular Disease. *Ann Intern Med.* 2021;174:1528-1541.
13. Fu EL, et al. Comparative effectiveness of SGLT2i versus GLP1-RA on cardiovascular outcomes in routine clinical practice. *Int J Cardiol.* 2022;352:172-179.